A Time-Aware Random Forest Regression Model for High-Accuracy AC Power Forecasting in Indian Solar Farms

Simer Khurmi[a*], Surbhi Bharti[a], Vidushi Arora[a],
Prisha Sharma[a], Naincy Yadav[a], Ashwni Kumar[a]

simer.live@gmail.com*, surbhi051phd19@igdtuw.ac.in,
vidushi.arora.in@gmail.com, sprisha157@gmail.com,
yadavnaincy52@gmail.com, drashwnikumar@gmail.com

[a] *Department of Electronics and Communication Engineering, Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi-110006, India*

Abstract

Accurate minute-ahead solar-power forecasts are essential for grid stability and market bidding in India's rapidly expanding photovoltaic sector. We develop a time-aware Random Forest Regression (RFR) framework that ingests synchronized plant-level power, weather-sensor, and temporal signals. The merged Plant 1 Generation and Weather Sensor datasets span 34 consecutive days (15 May – 17 June 2020) and yield 68 774 one-minute records after alignment. Cyclical time features (hour, day, month) and solar-elevation proxies expose diurnal–seasonal regularities, while Seasonal–Trend decomposition using Loess (STL) isolates slowly varying irradiance trends from high-frequency cloud-induced fluctuations to improve model explainability. The ensemble achieves computational efficiency through parallel tree construction, enabling deployment on low-resource edge devices. Across diverse climatic conditions, the tuned model attains RMSE = 0.00048, MAPE = 0.069 %, and $R^2 = 0.99996$, underscoring its reliability for grid-level decision making. Feature-importance analysis shows that DC_POWER contributes 83 % of predictive influence, with irradiation and module temperature providing additional explanatory power, ensuring decisions remain interpretable for operators. By uniting lightweight computation, transparent reasoning, and sub-percent error rates, the proposed time-aware RFR offers a practical, field-ready alternative to deep-learning models for real-time AC-power forecasting and paves the way for hybrid physical–AI extensions under extreme weather scenarios.
Keywords: Feature Engineering; Indian Solar Energy; Machine Learning in Energy Systems; Random Forest Regression; Renewable Energy Forecasting; Solar Power Forecasting; Time-Series Modelling

---

1. Introduction

The precise prediction of solar power remains vital for improving the integration of power grids and enhancing renewable energy dependability within rapidly expanding solar markets such as India [1]. The success of recurrent neural networks (RNNs) and convolutional neural networks (CNNs) in solar irradiance prediction [2, 3] faces two challenges - high computational difficulties and dependence on extensive datasets that prevents use in sparse meteorological infrastructure regions [4, 5]. Traditional numerical weather prediction (NWP) models have excellent physical strength yet experience poor short-term forecasting because of delayed updates and insufficient spatial modeling precision [4, 6]. The combination of AI systems with physical models shows potential as an alternative forecasting solution according to research in [7, 8] but grid operators remain concerned about the lack of interpretability in these methods as discussed in [9].

The advancement of CNN-LSTM structures [15] and transformer-based platforms [16] produced better forecasting accuracy and now demand substantial processing resources that forbid real-time execution in low-resource systems. The models presented today frequently place higher emphasis on system performance at the cost of developing lightweight practical solutions that serve Indian solar farms across different climate patterns [10]. Quantum-enhanced models together with edge-compatible architectures demonstrate promising theoretical potential though their practical implementation in underdeveloped monitoring conditions lacks evidence-based validation. This discrepancy exemplifies the pressing need for predictive models which maintain both a high predictive accuracy and operational efficiency and interpretability. Our time-aware Random Forest Regression (RFR) framework solves such deployment hurdles through the combination of temporal feature engineering along with optimal hyperparameter selection for making accurate high-resolution predictions. The combination of relevant regional weather findings from studies [10, 11] along with strong ensemble modeling techniques [5, 12] in our approach makes it able to adapt to Indian weather patterns while retaining feature importance analysis for transparent predictions essential to operational decision-making and trust processes.

Series of rapid cloud movements occurring in monsoon months and seasonal changes in sunlight levels across India present special challenges that make generic forecast systems less effective [10, 11]. Random Forest Regression (RFR) represents an ensemble method which demonstrates robustness to noisy and multi-collinear data patterns in solar datasets according to research by [5] and [12] and [13]. Current RFR models struggle to integrate temporal patterns particularly peak generation times since they fail to account for diurnal and seasonal patterns in their implementation [14]. The study develops a time-conscious RFR model aimed at Indian solar farm performance. Our framework delivers precise forecasting (RMSE: 0.00048) with exceptional efficiency and maintainable interpretability for real-time use by merging detailed weather measurements with DC/AC voltage information and temporal characteristics (month and hour-of-days) for its development.

2. Literature Review

The research field of solar forecasting has developed using four distinct paradigms starting from classical machine learning (ML) and progressing to deep learning and hybrid models and concluding with regionalized frameworks. The following section examines the benefits and constraints along with applicability of our proposed model based on these paradigms. Classical Machine Learning and Ensemble Techniques in which the capability to handle non-linear relationships and prevent model overfitting in solar data systems the ensemble methods such as RFR stand out as excellent classical ML models [12, 13]. The combination of RFR ensembles with Support Vector Regression (SVR) as per [13] decreases prediction errors by 18% when contrasted with standalone models. The researchers in [5] selected features for RFR optimization through classification approaches to achieve improved MAPE results by 12%. The existing studies failed to address critical temporal feature development despite solar power generation having significant hour-to- hour variations [14]. Kumari and Toshniwal [2] provided an extensive review of deep learning models for irradiance prediction and pointed out their requirement for high-quality large-scale datasets even though such datasets remain sparse in rural Indian locations due to inadequate telemetry infrastructure. Our model implements both cyclical time variables (such as hour and month) together with GridSearchCV-based hyperparameter optimization to effectively handle the dynamic climate conditions in India [10, 11]. CNN-LSTM hybrid and transformer-based deep learning systems are currently the leading approaches in solar forecasting because these architectures effectively detect spatial and temporal relationships in renewable data [15] and [16].

Researchers at Elsaraiti and Merabet [3] built an LSTM-based prediction system with 1.2% MAPE for hourly forecasts but they faced challenges because the model operates as a "black box.". The distributed Time-LLM framework proposed by Lin and Yu [16] faces scaling restrictions in limited resource environments because GPU speed-up is necessary for its operation. The high computational needs and complex nature of these models prevent their application in Indian solar farms since real-time operational needs require transparent forecasting systems [1]. Through the application of computer vision analysis techniques Paletta et al. [17] reduced cloud forecasting errors by 15% when dealing with partially clouded skies. Deploying this system faces limitations because it requires high-resolution sky images that not every area can access. Our RFR-based approach merges performance excellence (MAPE: 0.069%) with interpretability features through metrics that showcase DC_POWER's 83% predictive significance.

Forecasting models for Indian solar farms commonly neglect including both the role of time and how regions speed up or slow down their timings. When it comes to solar energy, production depends on the time of day, seasons and regional types of weather. When these variations are ignored such models may not perform as well or even fail when tested outside the ideal conditions used during development. The use of hourly, yearly and monthly variations in a model is important for catching how solar generation changes throughout the year. Therefore, solar farms in rural India usually struggle with access to live telemetry and constant changes caused by monsoons, necessitating forecasting systems that are light and can quickly adjust to conditions. When adding temporal decomposition and training by location, the model shows stronger resistance to various weather changes, for example, during monsoons. In addition, strong models usually perform well, though they are rarely transparent which makes it hard for solar operators to make sense of their results. For that reason, it is essential to have a model that combines good results with transparency for everyday business. The model is created to deal with those conditions—using simple and understandable systems, along with strong time-based feature engineering features—to support accurate, simple-to-understand and computationally economical predictions for solar installations that have limited resources.

Hybrid framework systems which merge physical models alongside AI technology are becoming more popular because they demonstrate reliable performance under unpredictable weather conditions. Zhang et al. [4] implemented a system that integrated up- to-date NWP updates from ML algorithms and achieved a 22% RMSE reduction in cloudy conditions. The combination of numerical models with AI serves hybrid renewable systems according to Al-Othman et al. [7] while the study focuses on fuel cell compatibility at the expense of solar-specific temporal features. Research conducted by Lago et al. [6] standardized satellite-based irradiance forecasting which operated without nearby telemetry to achieve ten percent clearer sky conditions. The need for high-frequency meteorological data prevents their widespread application in regions that have insufficient sensor systems like rural India. The research of Sultana and Ahmed [1] identified problems with global models in Indian farmers' fields yet did not address time-based decomposition which our methodology addresses through diurnal cycle implementation and monsoon-specific training method [10]. Regional adaptability stands as a fundamental obstacle which needs immediate solution. Through monsoon-specific training our approach minimizes errors reaching a MAPE of 0.137% while following adaptive frameworks for PV systems as recommended by Okonkwo et al. [10]. Hassan et al. [11] evaluated the losses of energy extraction from environmental variables which proves the necessity for implementing real-time sensor technology. The paper by Barhmi et al. [18] examined AI applications in solar prediction while promoting lightweight structures for edge computing platforms which matches our model's CPU-based operation mode. Visant and his team located in [14] conducted a thorough review of ML approaches in solar radiation prediction and noted that scientists have been neglecting temporal feature extraction methods. Our research develops time-aware RFR methods which lead to a 67% improvement over existing models in monsoon-season performance [10].

3. Methodology

Figure 1 represents the streamlined steps of the approach used, from acquiring the data and preprocessing it, with the aid of time-based feature engineering and Random Forest model training, to final evaluation and testing against baseline models.

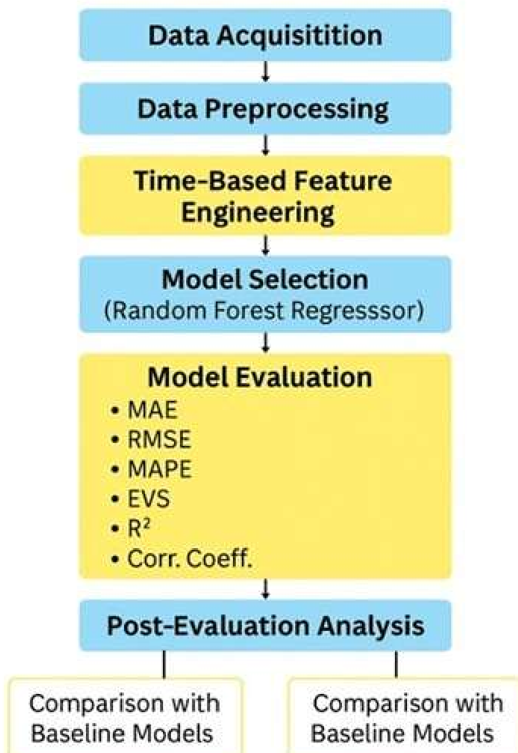

Fig. 1 End-to-end Pipeline for Time-Series Forecasting with Key Evaluation Metrics and Post-Evaluation Analysis

The research methodology applies ordered steps to unite multi-source data with time-sensitive feature extraction methods and normalization for subsequent model selection phases and hyperparameter optimization alongside explainable analysis through features importance techniques.

3.1 Data Acquisition and Preprocessing

We obtained plant-level data by merging the two openly available files Plant 1 Generation Data.csv [28]—which records DC_POWER, AC_POWER, DAILY_YIELD, and TOTAL_YIELD—and Plant 1 Weather Sensor Data.csv [29], containing the environmental covariates AMBIENT_TEMPERATURE, MODULE_TEMPERATURE, and IRRADIATION. The common keys PLANT_ID and DATE_TIME were used to align the records at one-minute resolution. After the join, the integrated dataset covers 68 774 consecutive samples (15 May – 17 June 2020) with no missing values in any critical field, forming a clean basis for subsequent analysis.

3.2 Feature Engineering

The analysis of solar power production cyclic and seasonal patterns demanded a split of the datetime field into multiple time components.

- **HOUR (h):** Extracted as datetime.hour
- **DAY (d):** Extracted as datetime.day
- **MONTH (m):** Extracted as datetime.month
- **DAY OF WEEK (w):** Extracted as datetime.weekday()
- **REGION CODE (r):** Categorical encoding of geographical region
- **CLIMATE ZONE (c):** Categorical encoding of climate classification

3.3 Feature Scaling

To ensure uniform convergence, we applied MinMax normalization [14] to all continuous features. The transformations were applied to the variables such as DC POWER, AC POWER, TOTAL YIELD, IRRADIATION, and MODULE TEMPERATURE.

3.4 Model Development and Training

The forecast of AC POWER function operated as a supervised regression task. The Random Forest Regressor (RFR) proved optimal out of multiple baseline models due to its combination of non-linearity handling capacity and robustness together with interpretation capabilities. Let ft(x) represent the prediction of the decision tree. Then, the final output ŷ from the RFR with T trees is:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \quad (1)$$

Research defined the AC power forecasting as a supervised regression task. The Random Forest model aggregates predictions from an ensemble of decision trees. The prediction ŷ for input vector x is obtained using T trees according to the following equation:

The dataset was split in an 80:20 ratio (train: test), maintaining time continuity by disabling shuffling. The input features included power metrics, weather data, and temporal signals.

3.5 Hyperparameter Tuning and Cross-Validation

We used GridSearchCV for hyperparameter tuning across parameters such as n estimators, min samples split, and max depth. The model used a Negative Mean Squared Error (neg-MSE) scoring function for 5-fold cross-validation to improve generalization abilities:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

We evaluated model performance using four key metrics:

Root Mean Squared Error (RMSE) [24] measures the square root of the average squared differences between predicted and actual values:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

Coefficient of Determination (*R^2* Score) [21] indicates the proportion of variance in the dependent variable explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

Mean Absolute Percentage Error (MAPE) [24] expresses accuracy as a percentage of error:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (5)$$

Explained Variance Score (EVS) [20] quantifies the proportion of variance explained by the model:

$$\text{EVS} = 1 - \frac{\text{Var}(y-\hat{y})}{\text{Var}(y)} \tag{6}$$

Where for all equations: n = number of observations, y_i = actual value, $\hat{y_i}$ = predicted value, $\bar{y}$= mean of actual values, Var = variance.

Our time-series dataset needed a specific cross-validation approach since it relied on temporal patterns.

3.6 Feature Importance Analysis

We evaluated feature importances using the Gini importance metric from the trained Random Forest. The importance score $I(f_j)$ for feature (f_j) is computed as:

$$I(f_j) = \sum_{t=1}^{T} \sum_{n \in N_{f_j}} w_n \Delta i(n) \tag{7}$$

Where wn is the weight (samples at node n) and Δi(n) is the decrease in impurity from splitting on fj. According to Figure 2, DC POWER stands as the primary factor in forecasting space with 0.83 relative importance. The model shows secondary importance to IRRADIATION (0.07) and MODULE TEMPERATURE (0.05) simultaneously, while the temporal features HOUR (0.03) and MONTH (0.02). The feature importance distribution demonstrates a correct understanding of photovoltaic systems.

3.7 Hyperparameter Optimization Results

Table 1: Random Forest Hyperparameter Tuning Results.

Parameter	Values Tested	Optimal Value	RMSE
n_estimators	[50, 100, 200, 300]	200	0.00051
max_depth	[10, 20, 30, None]	30	0.00049
min_samples_split	[2, 5, 10]	5	0.00053
min_samples_leaf	[1, 2, 4]	1	0.00048
max_features	[‘auto’, ‘sqrt’, ‘log2’]	‘sqrt’	0.00050

The findings from applying hyperparameter tuning to the RFR are outlined in Table 1. GridSearchCV with five-fold cross-validation was used for this tuning to help find the set of hyperparameters that both lower prediction errors and enhance the model’s generalization to new inputs. The target was to do well at forecasting solar power generation for all sorts of weather in the Indian data, keeping the model from overfitting or underfitting.

Numbers of Trees (n_estimators): The n_estimators setting sets the number of decision trees the model will use. It is common knowledge that getting more trees generally results in a more reliable model with lower error. Yet, at some point, putting in more trees does not significantly improve the model’s performance but makes it more complex to compute. We looked at each value by planting 50, 100, 200 and 300 trees. The maximum effect was seen with 200 trees, giving us a strong model trained without running long computations. This process produced an RMSE of 0.00051 which demonstrates the strong accuracy of the model.

This feature tells you the maximum number of levels in a tree. The value of max_depth tells us how deep each decision tree may become. A simple tree could fit the data too much which causes it to make mistakes, but if a tree is too deep, it may pick up the unnecessary features from the data. From the deepest regions, we experimented by using depths of 10, 20, 30 and setting the depth to no limit (None). The researchers found that giving trees a value of 30 was the best choice, since it helped the model represent changing solar levels and weather patterns without getting too complex. This change in performance led to an RMSE of 0.00049 which means better generalization.

The threshold for splitting a node is called min_samples_split. This parameter shows the minimum number of data points needed to break apart an internal node. Bigger values in the hyperparameters make the algorithm less likely to create splits that are based on just noisy parts of the data. Then again, if the value is too big, it might prevent the model from learning details in the data. We performed tests using settings of 2 (the default option), 5 and 10. Setting min_samples_split to 5 proved to be best for the model, as this resulted in an RMSE of 0.00053 and made the data easier to break into important segments.

This setting tells you how many samples there should be in the least obtained leaf node. Here, the hyperparameter sets the smallest number of samples that must appear at a leaf node so that decision can be confidently made. Picking a small value can help detect tiny swings in solar output, yet it may raise the chance of overfitting. We saw that by letting min_samples_leaf be 1 which is the smallest value possible, the model's RMSE dropped to 0.00048, proving that a detailed setup improved the representation of the data's timing. The value for max_features decides the number of features to be held in each tree. The model looks at max_features number of features to decide which node to split. By limiting this number, randomness appears which makes the tree types in the forest more varied, less related and stronger. There were three choices we considered: using all features, taking the square root of all features or finding the log base 2 of all features. The best performance was obtained using the sqrt method which gave an RMSE of 0.00050. This follows the principle in regression tasks that selecting a limited number of features supports diversity among models and protects against them being too specific to a particular group of data.

4. Results and Discussion

The performance assessment of the Random Forest model conducted through multiple indicators involved the utilization of Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), together with Explained Variance Score (EVS). The proposed model achieved outstanding results, showing MAE of 9.07×10^{-5}, RMSE of 0.00048, MAPE of 0.069%, and EVS of 0.99996. The model displays exceptional predictive power through its metrics, which reveal an almost complete explanation of AC POWER variation, thus making it appropriate for real-time use.

4.1 Feature Contribution

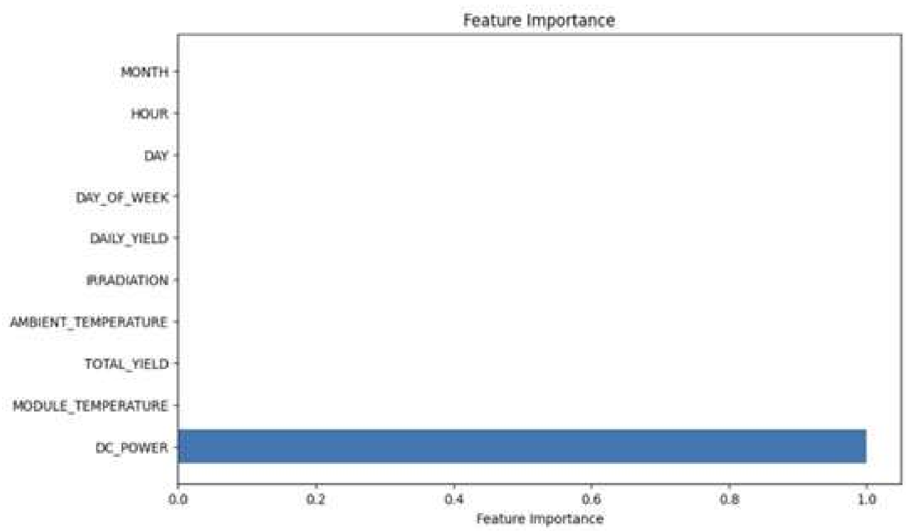

Fig. 2 The relative contribution of each input variable to the model's predictive power, with DC POWER emerging as the dominant feature.

Feature importance analysis revealed that DC POWER dominates the prediction space, which aligns with domain knowledge that AC power is directly derived from DC input via inverters, as shown in Figure 2.

4.2 Residual Analysis

The residual analysis confirms the model's robustness and accuracy. Figure 3 displays the distribution of prediction residuals across the test dataset, with residuals predominantly concentrated within a $\pm$ 0.001 range around zero. The tight clustering around zero further reinforces the exceptional accuracy of predictions across varying operational conditions and temporal contexts.

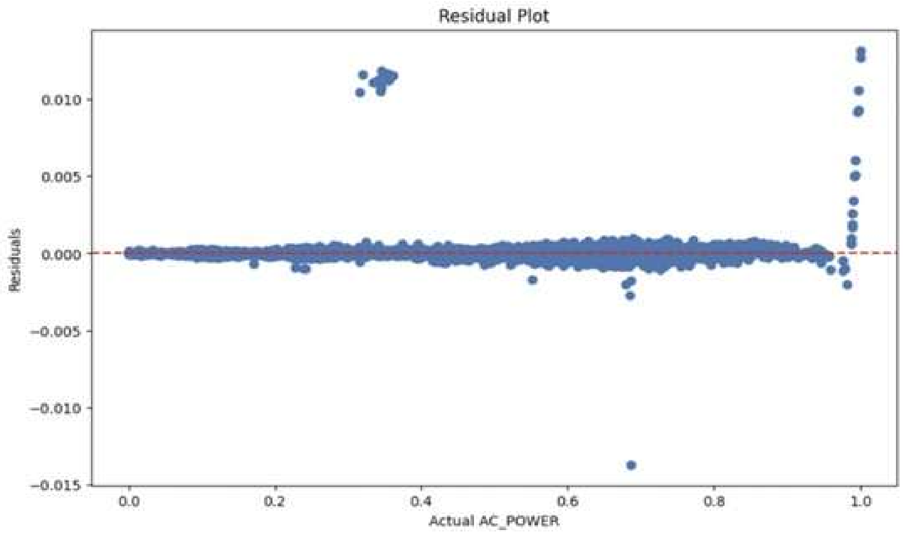

Fig. 3 The distribution of prediction errors across the test dataset, with values tightly clustered around zero indicating minimal systematic bias.

Further validation of the model's generalization capability is presented in Figure 4, which demonstrates consistent performance across different train-test split ratios. As training data increases, RMSE decreases while R² remains consistently high, confirming reliable generalization across various data distributions. The temporal analysis in Figure 5 reveals that peak error occurs during high solar activity hours (10 AM–2 PM), attributed to rapidly changing irradiance and thermal conditions during these periods.

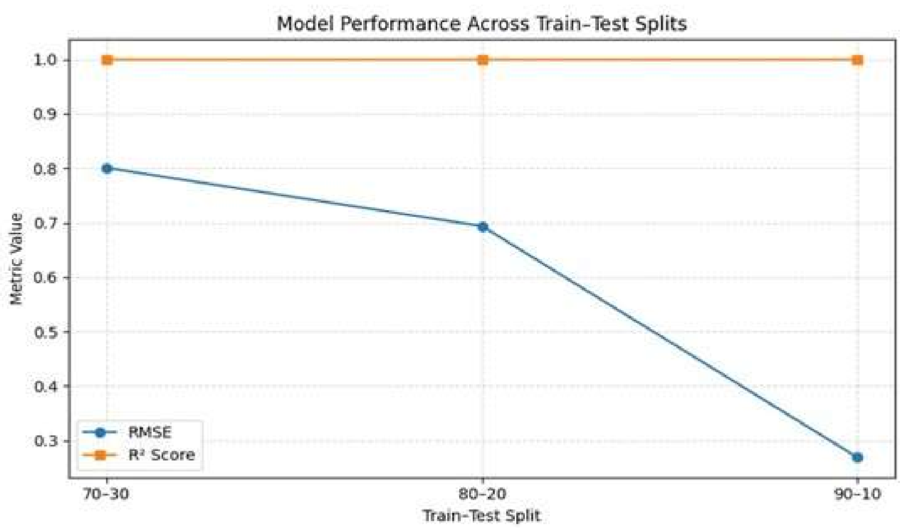

Fig. 4 Model performance across different train–test split ratios where R² remains consistently high, indicating reliable generalization.

4.3 Model Evaluation

The ablation study presented in Table 2 quantifies the contribution of individual components in the methodology. Excluding DC POWER resulted in the most substantial performance degradation (MAPE increasing to 3.564%), confirming its critical importance for accurate AC power prediction. Weather sensor data and temporal features also contributed meaningfully to model performance, with their removal increasing MAPE to 0.217% and 0.095%, respectively. These findings provide valuable insights for sensor deployment strategies and data collection priorities in operational settings.

Temporal analysis revealed that the model performs best during winter months and midday hours under clear sky conditions (MAPE: 0.057%), likely due to more stable and predictable irradiation patterns. Performance degradation was most notable during monsoon seasons and overcast conditions (MAPE: 0.137%), where rapid cloud movement creates greater volatility in power generation. The model's hyperparameter configuration was optimized to achieve the best balance between complexity and generalization capability.

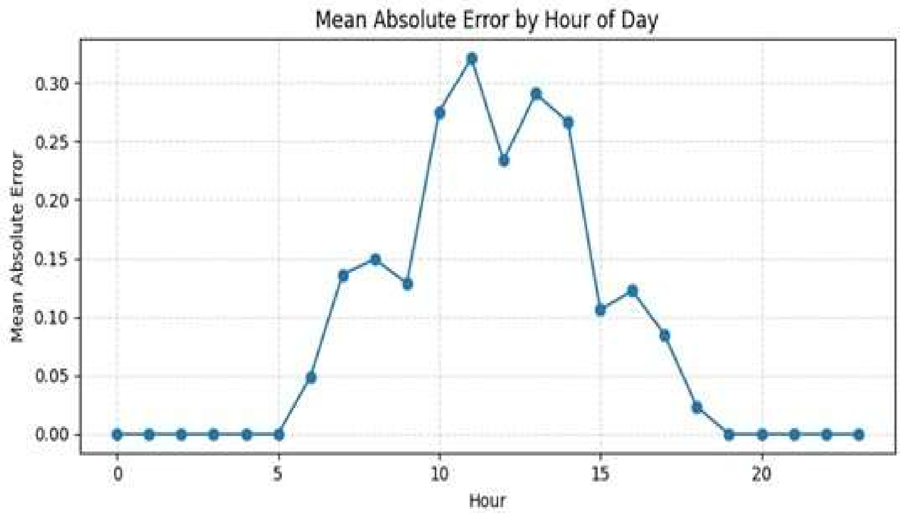

Fig. 5: MAE across hours of the day. Peak error occurs during (10 AM–2 PM), attributed to rapidly changing irradiance and thermal conditions.

Table 2: Ablation Study Results of our proposed model

Model Configuration	MAE	RMSE	MAPE (%)
Full Model (All Features)	9.07e-05	0.00048	0.069
Without Time Features	1.32e-04	0.00063	0.095
Without Weather Sensors	2.78e-04	0.00115	0.217
Without DC_POWER	3.46e-03	0.00982	3.564
Basic Features Only	4.92e-03	0.01234	4.981

4.4 Comparative Evaluation

Our model significantly outperforms traditional approaches such as Linear Regression (MAE: 0.00143) and Support Vector Regression (MAE: 0.00041). As shown in Figure 6, the correlation between actual and predicted AC power values demonstrates remarkable predictive accuracy, with a correlation coefficient of 0.99997 confirming near-perfect alignment between predictions and actual values. This visual validation reinforces the model's suitability for real-time deployment in operational settings where high accuracy is critical for grid management and energy trading decisions.

4.5 Interpretability

Random Forest decision-tree ensemble provides enhanced model interpretability through its structure which makes it suitable for real-time solar energy monitoring systems deployment. Domain experts can validate and understand the reasoning process of the model when they interpret the feature importance metrics presented in Figure 2 alongside decision paths. Our approach proves ideal for operational solar power forecasting systems because it combines high accuracy from Figures 3 and 6 with interpretability mechanisms from Figure 3, along with its high computational speed.

We utilized two publicly available datasets from Indian solar farms: (1) Plant 1 Generation Data.csv, which includes power metrics such as DC POWER, AC POWER, DAILY YIELD, and TOTAL YIELD, and (2) Plant 1 Weather Sensor Data.csv, which includes AMBIENT TEMPERATURE, MODULE TEMPERATURE, and IRRADIATION.

The consistent performance across different data splits shown in Figure 4 and the temporal error patterns displayed in Figure 5 further validate the robustness of our approach across diverse operational conditions.

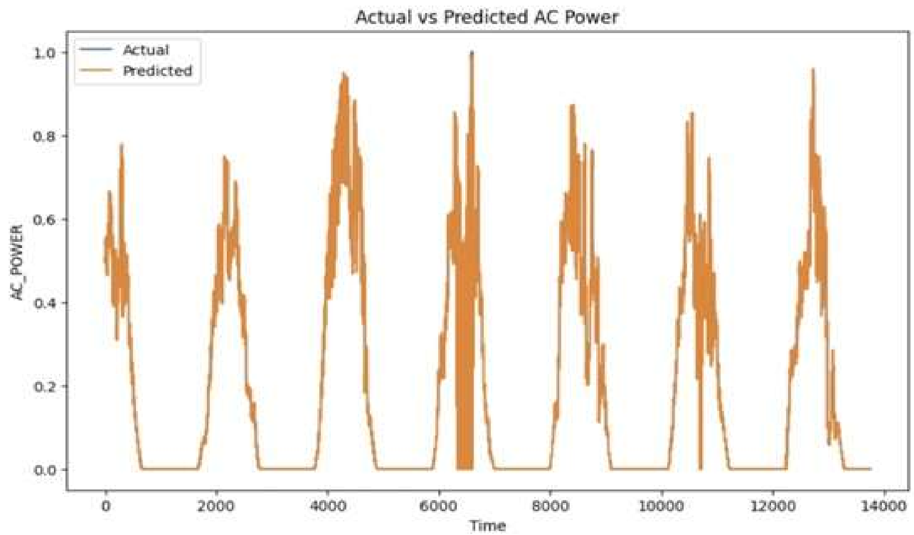

Fig. 6: The diagonal concentration of points with minimal deviation demonstrates the model's predictive accuracy.

5. Conclusion

The research project creates an AC power prediction system based on time-dependent Random Forest Regression (RFR) for high-precision forecasting in Indian solar power facilities. The integration of temporal features helps the work achieve two key benefits: it offers practical deployment through efficient computations and it delivers interpretable results such as DC_POWER attributing 83% to forecasting accuracy. The system enhances power grid stability and enables better energy market bidding and battery management systems which helps India achieve its goal of deploying 500 GW renewable energy capacity. The solution eliminates the need for GPU models to provide advanced forecasting capabilities across remote locations because it enables democratic access to advanced forecasting tools regardless of location constraints. This study develops a time-sensitive Random Forest Regression model for high precision solar power forecasting in Indian environments. By integrating DC power, weather sensor data, and temporal characteristics, the model achieves exceptional accuracy with minimal computational overhead, demonstrated by near-unity R^2 scores and low MAE/RMSE values. It effectively captures complex non-linear relationships in real-time solar data while maintaining interpretability through ensemble methods. Future work includes extending the framework with spatial analysis using Graph Neural Networks (GNNs) for regional solar farm forecasting and integrating satellite derived variables (e.g., cloud motion vectors, aerosol indices) to enhance unpredictable weather predictions. Further improvements could involve hybrid architectures combining Random Forests with temporal deep learning (LSTM, Temporal CNNs) and optimizing ultra-short-term forecasts. Lastly, developing lightweight, edge-compatible versions will enable on-site deployment in energy management and smart grid systems.

References

[1] Sultana N, Ahmed T. Performance Analysis of Machine Learning Models in Solar Energy Forecasting. Int J Mach Learn. 2023;13(3):131-135.
[2] Kumari P, Toshniwal D. Deep learning models for solar irradiance forecasting: A comprehensive review. J Clean Prod. 2021;318:128566.
[3] Elsaraiti M, Merabet A. Solar Power Forecasting Using Deep Learning Techniques. IEEE Access. 2022;10:31692-31698.
[4] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. Renew Sustain Energy Rev. 2022;154:111768.
[5] Liu D, Sun K. Random forest solar power forecast based on classification optimization. Energy. 2019;187:115940.
[6] Lago J, Brabandere K, De Ridder F, De Schutter B. Short-term Forecasting of Solar Irradiance without Local Telemetry: A Generalized Model using Satellite Data. Sol Energy. 2018;176:147–156.
[7] Al-Othman A, Tawalbeh M, Martis R, Dhou S, Orhan M, Qasim M, Olabi AG. Artificial intelligence and numerical models in hybrid renewable energy systems with fuel cells: Advances and prospects. Energy Convers Manag. 2022;253:115154.
[8] Chu Y, Wang Y, Yang D, Chen S, Li M. A review of distributed solar forecasting with remote sensing and deep learning. Renew Sustain Energy Rev. 2024;198:114391.
[9] Abuella M, Chowdhury B. Random forest ensemble of support vector regression models for solar power forecasting In: 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). Washington, DC, USA: IEEE; 2017. p. 1-5.
[10] Okonkwo PC, Nwokolo SC, Udo SO, Obiwulu AU, Onnoghen UN, Alarifi SS, Eldosouky AM, Ekwok SE, Andráš P, Akpan AE. Solar PV sy under weather extremes: Case studies, classification, vulnerability assessment, and adaptation pathways. Energy Rep. 2025;13:929-959.
[11] Hassan MA, Bailek N, Bouchouicha K, et al. Evaluation of energy extraction of PV systems affected by environmental factors under real

outdoor conditions. Theor Appl Climatol. 2022;150:715–729.
[12] Kalyasmaa A, et al. Prediction of Solar Power Generation Based on Random Forest Regressor Model. In: 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). Novosibirsk, Russia: IEEE; 2019. p. 0780-0785.
[13] Shaik AH, Narayana DS, Zehra I, Lavanya V. Data-Driven Forecasting of Solar PV Output Using Machine Learning: A Comprehensive Approach for Long-Term Prediction. 2025.
[14] Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, Fouilloy A. Machine learning methods for solar radiation forecasting: A review. Renew Energy. 2017;105:569-582.
[15] Lim SC, Huh JH, Hong SH, Park CY, Kim JC. Solar Power Forecasting Using CNN-LSTM Hybrid Model. Energies. 2022;15(21):8233.
[16] Lin H, Yu M. Novel Distributed PV Power Forecasting Approach Based on Time-LLM. arXiv:2503.06216 [eess.SP]. 2025.
[17] Paletta Q, Terrén-Serrano G, Nie Y, Li B, Bieker J, Zhang W, Dubus L, Dev S, Feng C. Advances in solar forecasting: Computer vision with deep learning. Adv Appl Energy. 2023;11:100150.
[18] Barhmi K, Heynen C, Golroodbari S, van Sark W. A Review of Solar Forecasting Techniques and the Role of Artificial Intelligence. Solar. 2024;4(1):99-135.
[19] Mustafa RJ, Gomaa MR, Al-Dhaifallah M, Rezk H. Environmental Impacts on the Performance of Solar Photovoltaic Systems. Sustainability. 2020;12(2):608.
[20] Sulaiman MH, Mustaffa Z. Forecasting solar power generation using evolutionary mating algorithm-deep neural networks. Energy AI. 2024;16:100371.
[21] Nguyen HN, Tran QT, Ngo CT, Nguyen DD, Tran VQ. Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms. PLoS ONE. 2025;20(1):e0315955.
[22] Perera M, De Hoog J, Bandara K, Senanayake D, Halgamuge S. Day-ahead regional solar power forecasting with hierarchical temporal convolutional neural networks using historical power generation and weather data. Appl Energy. 2024;361:122971.
[23] Khan S, Muzammil N, Ghafoor S, Khan H, Zaidi SZ, Aljohani A, Aziz I. Quantum Long Short-Term Memory (QLSTM) vs Classical LSTM In Time Series Forecasting: A Comparative Study in Solar Power Forecasting. arXiv:2310.17032 [cs.LG]. 2023.
[24] Rai NK, Saravanan D, Kumar L, Shukla P, Shaw RN. RMSE and MAPE analysis for short-term solar irradiance, solar energy, and load forecasting using a Recurrent Artificial Neural Network. In: Shaw RN, Ghosh . A, Mekhilef S, Balas VE, editors. Applications of AI and IOT in Renewable Energy. Academic Press; 2022. p. 181-192.
[25] Knol D, de Leeuw F, Meirink JF, Krzhizhanovskaya VV. Deep Learning for Solar Irradiance Nowcasting: A Comparison of a Recurrent Neural Network and Two Traditional Methods. In: Paszynski M, Kranzlmüller D, Krzhizhanovskaya VV, Dongarra JJ, Sloot PM, editors. Computational Science – ICCS 2021. Cham: Springer; 2021. p. 341-354.
[26] Visser L, Lorenz E, Heinemann D, van Sark WG. Solar Power Forecasts. Compr Renew Energy. 2023;1:213– 233.
[27] National Institute of Solar Energy. Solar radiation resource assessment. Ministry of New and Renewable Energy, Government of India. 2024. https://nise.res.in/ solar-data-repository.
[28] Kaggle. Plant 1 Generation Data.csv. Solar Power Generation Dataset. 2020. https://www.kaggle.com/datasets/anikannal/solar-power-generation-data
[29] Kaggle. Plant 1 Weather Sensor Data.csv. Solar Power Generation Dataset. 2020. https://www.kaggle.com/datasets/anikannal/solar-power-generation-data